CIAOER RP 75-08 OUO 1 of 1

A Model for Forecasting Soviet Grain Yields Mar 75 ER RP 75-08

For Official Use Only

A Model for Forecasting Soviet Grain Yields

ER RP 75-8
March 1975

Copy N° 47

TABLE OF CONTENTS

TABLES

MAP

A MODEL FOR FORECASTING SOVIET GRAIN YIELDS

INTRODUCTION

1. For many years, this Office has assessed the progress of the Soviet grain crop during the growing season. Independent assessments are necessary mainly because the USSR Ministry of Agriculture provides only the most general situation reports during the growing season. Crop results are reported several months after the completion of the harvest. The tautness of the world food supply and the consequences of the wide swings in Soviet grain production have increased the need for timely forecasts of Soviet grain output.

2. This publication documents the development of a model to predict grain yields in 33 crop regions (see the map). The predicted yields—based on time trends and a composite index of several weather variables—are combined with reported data on sown area to obtain crop estimates. With the help of the model, crop estimates can be made as early as April and revised to take account of additional information until the harvest is completed.

3. The model's structure, computer programs, and weather data base can be applied to any crop for which yield data are available. So far, the model has been used to forecast yields for all grain, winter wheat, and spring wheat.[1] To estimate the Soviet grain crop, estimation of sown area and adjustments to reflect collateral information are also used.

PRINCIPAL FINDINGS

4. A weather-yield model has proved to be useful in estimating Soviet crop yields. The predicted total 1974 grain harvest was 2% above the actual harvest. The model can produce reasonably reliable estimates early in the season and can then be revised every ten days as new weather data are received.[2]

5. Data on production, sown area, and yields in 1958-73 have been collected for all grain, winter wheat, spring wheat, winter rye, and spring barley. A weather data base covering the period from 1961 to the present has also been assembled for use with the model. Relying on this data base, the model employs a weather index—representing the influence of temperature and precipitation—and a time trend—representing technological change—as explanatory variables in a set of independent linear equations that predict crop yields.

NOTE: Comments and queries regarding this publication are welcomed. They may be directed to [redacted]f the Office of Economic Research, Code 143, Extension 6811.

STATINTL

[1] In this publication, the official Soviet definition of all grain is used—including wheat, rye, barley, oats, corn, rice, millet, buckwheat, and pulses.

[2] This model was developed in early 1973 and formed the basis of CIA estimates for that year. The model was modified slightly in 1974. This publication is limited to the current model, and all data refer to it unless otherwise stated.

6. The 1974 [illegible] The USSR has announced a total grain harvest of 195.5 million metric tons and two republic harvests—111.6 million tons for the RSFSR and 46 million tons for the Ukraine. The model's final prediction (end of August) for the USSR was 199.4 million tons, within 2% of the total announced in December.[a] In September, Politburo Chairman Brezhnev said the crop would be "not bad" but admitted that areas in Siberia and Kazakhstan were having trouble. The announcement of a crop of 195.5 million tons confirms these indications.

7. The prediction for 1974 also seems reasonable in terms of its development over time and its forecast of regional trends. The prediction increased sharply during May and June, when the weather was favorable, and then declined as conditions deteriorated. The final predictions for the RSFSR and the Ukraine (115.3 million tons and 47.9 millions tons, respectively) are slightly higher than the published totals (3% and 4%). The prediction for the remaining portion of the USSR is 4% below the actual harvest.

8. Although the model describes the historical data quite well, prediction errors can be significant. Tests show that an error of 2% to 8% in the predicted national yield can be expected. An error in estimates of the sown area can either compound or mitigate the effect of errors in predicted yields.

FORMULATION OF THE MODEL

Factors Influencing Yields

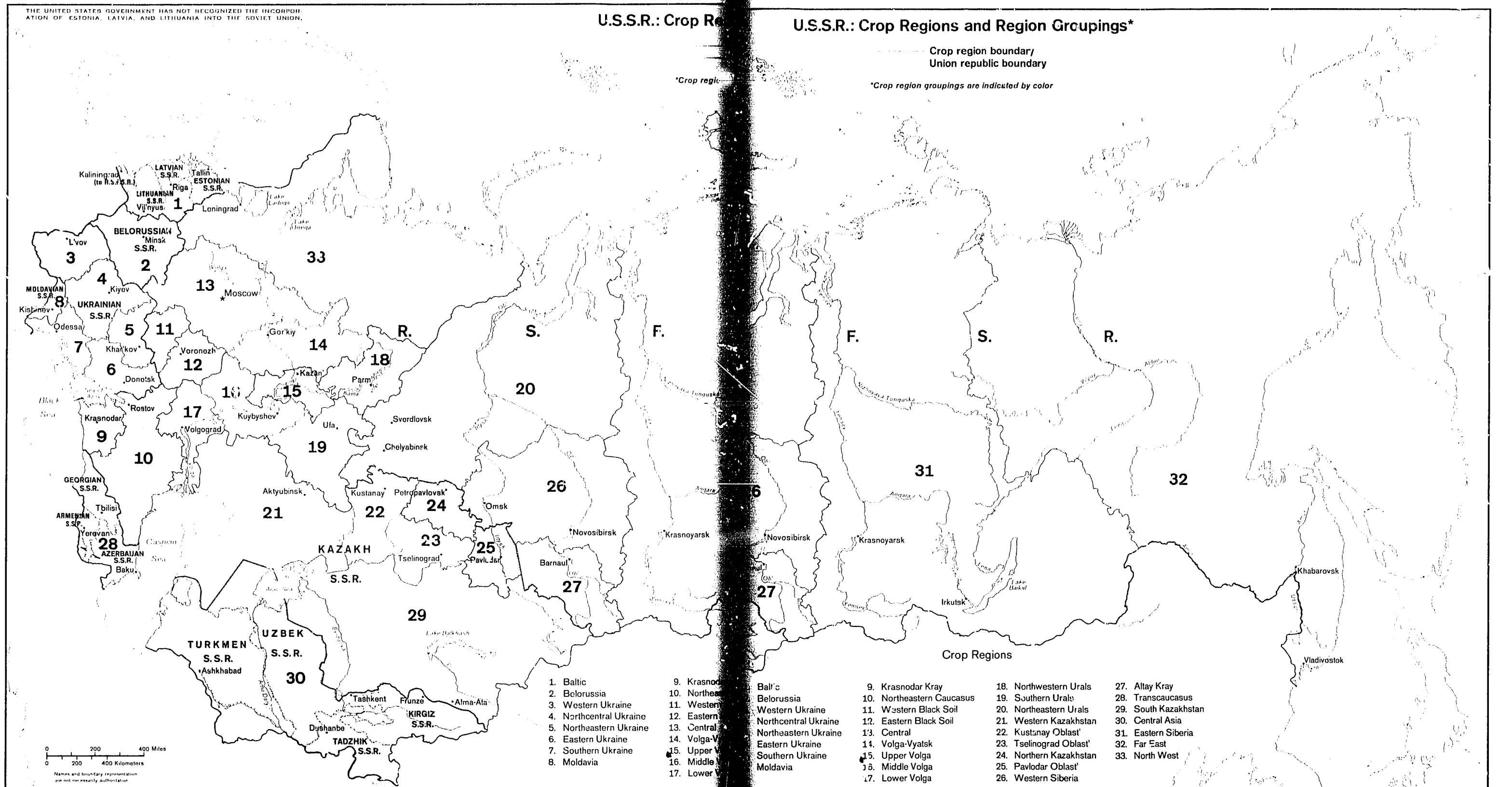
Weather

9. Although the importance of weather in determining grain yields is apparent, the causal relationships are difficult to define. Many individual weather variables affect the final yield. Moisture, for example, influences the number of plants per hectare, the number of stalks per plant, leaf area, the number of heads per plant, the number of kernels per head, and the weight per kernel. Temperature variations speed up or curb growth and change the water balance in the plant. Extreme temperatures may injure or kill the plant.

10. Other weather variables that influence yields include sunlight, humidity, hail, and wind. In particular, the hot *sukhovey* (dry winds) prevalent in the eastern grain-growing regions of the USSR greatly increase evaporation and transpiration, and at times even beat down a crop. Weather variables also affect yields indirectly because of the importance of good weather in fieldwork at planting and harvest times and because of their role in encouraging or retarding plant diseases, parasites, and weeds.

11. Specific patterns of interaction between weather variables and the physiological requirements of plants (and hence yields) vary considerably according to the peculiarities of local climates. Nonetheless, certain common factors are important. In the early stages of growth—at least through tillering—grain requires a continuous supply of moisture in the upper layer of the soil. After tillering, which normally occurs a little less than 30 days after sowing, the period of rapid vegetative growth occurs. As stems and leaves develop, consumption of water by the plants increases greatly. The heading stage, which is critical to grain yields, usually

[a] The predictions represent gross grain production—the grain obtained from the harvesting machine in the field, including excess moisture, unripe and damaged kernels, weed seeds, and the losses in handling and transporting the grain between the field and storage facilities.


THE UNITED STATES GOVERNMENT HAS NOT RECOGNIZED THE INCORPORATION OF ESTONIA, LATVIA, AND LITHUANIA INTO THE SOVIET UNION.
U.S.S.R.: Crop Regions and Region Groupings*
Crop region boundary
Union republic boundary
*Crop region groupings are indicated by color
Kaliningrad (to R.S.F.S.R.)
LATVIAN S.S.R.
Riga
Tallin
ESTONIAN S.S.R.
LITHUANIAN S.S.R.
Vil'nyus
Leningrad
BELORUSSIAN
Minsk
S.S.R.
L'vov
Kiyev
MOLDAVIAN S.S.R.
Kishinev
UKRAINIAN S.S.R.
Odessa
Khar'kov
Donetsk
Voronezh
Moscow
Gor'kiy
Kazan'
Perm'
Sverdlovsk
Chelyabinsk
Ufa
Kuybyshev
Rostov
Volgograd
Krasnodar
Black Sea
GEORGIAN S.S.R.
Tbilisi
ARMENIAN S.S.R.
Yerevan
AZERBAIJAN S.S.R.
Baku
Caspian Sea
Aktyubinsk
Kustanay
Petropavlovsk
Omsk
Tselinograd
Pavlodar
KAZAKH
S.S.R.
Novosibirsk
Barnaul
Krasnoyarsk
Irkutsk
Khabarovsk
Vladivostok
TURKMEN S.S.R.
Ashkhabad
UZBEK S.S.R.
Tashkent
Frunze
Alma-Ata
KIRGIZ S.S.R.
Dushanbe
TADZHIK S.S.R.
R.
S.
F.
S.
R.
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
0 200 400 Miles
0 200 400 Kilometers
Names and boundary representation are not necessarily authoritative
Crop Regions
1. Baltic
2. Belorussia
3. Western Ukraine
4. Northcentral Ukraine
5. Northeastern Ukraine
6. Eastern Ukraine
7. Southern Ukraine
8. Moldavia
9. Krasnodar Kray
10. Northeastern Caucasus
11. Western Black Soil
12. Eastern Black Soil
13. Central
14. Volga-Vyatsk
15. Upper Volga
16. Middle Volga
17. Lower Volga
18. Northwestern Urals
19. Southern Urals
20. Northeastern Urals
21. Western Kazakhstan
22. Kustanay Oblast'
23. Tselinograd Oblast'
24. Northern Kazakhstan
25. Pavlodar Oblast'
26. Western Siberia
27. Altay Kray
28. Transcaucasus
29. South Kazakhstan
30. Central Asia
31. Eastern Siberia
32. Far East
33. North West
502400 1-75

occurs a little more than 30 days after tillering.[4] The relatively high temperatures that generally prevail in the USSR during this stage result in increased transpiration and in the more rapid depletion of soil moisture by evaporation. At that stage, therefore, the plants require more rainfall than during earlier stages of development.

12. After heading, the plant's dependence on moisture and its sensitivity to temperature decrease. Excessive rainfall still may damage the crop, however, by causing lodging (matting) and by promoting diseases—such as rust, scab, mildew, and leaf spot—as well as weed growth. As in the earlier stages of growth, excessive temperatures may be injurious, especially if accompanied by dry winds. Too low a temperature preceding the harvest may delay ripening so that the crop is caught by early frost.

13. Links between weather factors and the stages of growth cannot be defined with precision because of annual variations in sowing dates and in the seasonal pattern of growth among different grains and areas. For example, the same grain in the same area may be planted at substantially different times in different years, and separate varieties planted in the same region may grow at different rates.

Technology

14. Crop yields have increased steadily since 1958 in many regions of the USSR but very slowly in other regions (see Table 1).[5] The systematic improvement that has occurred probably reflects increased use of agricultural chemicals—fertilizer, lime, and insecticides—and better and more timely cultivation and harvesting. All of these changes are technological improvements. Unfortunately,

---

[4] These stages of development refer to spring-sown small grains. Fall-sown winter grains usually tiller before entering dormancy.

[5] Data for earlier years could not be used, because in 1958 the official measure of the grain crop is believed to have changed over to a new basis. For a discussion of this question, see Eberhard Schinke, "Soviet Agricultural Statistics" in Vladimir G. Treml and John P. Hardt (eds.), *Soviet Economic Statistics*, p. 242.

Table 1

USSR: Grain Yields

Centners per Hectare

Year	All Grain				Winter Wheat			Spring Wheat		
	USSR	RSFSR	Ukraine	Kazakhstan	USSR	RSFSR	Ukraine	USSR	RSFSR	Kazakhstan
1958	11.1	10.6	16.8	9.5	16.2	17.5	17.6	9.7	9.9	9.6
1959	10.4	9.9	16.1	8.7	15.2	13.5	18.6	9.4	9.8	8.8
1960	10.9	10.7	15.8	8.5	15.1	16.1	17.5	9.5	10.2	8.4
1961	10.7	9.9	19.9	6.6	16.9	15.6	21.9	8.2	9.0	6.9
1962	10.9	11.0	17.9	6.5	16.8	18.6	17.6	8.2	9.2	6.5
1963	8.3	8.3	12.9	4.4	12.9	13.4	14.2	5.9	7.0	3.8
1964	11.4	10.7	17.7	9.8	13.8	12.9	17.0	9.9	9.9	9.9
1965	9.5	9.0	19.2	3.1	16.1	14.1	21.3	5.5	6.8	3.1
1966	13.7	13.1	21.5	10.8	20.4	19.6	24.8	12.0	12.6	11.0
1967	12.1	11.9	20.5	6.3	17.8	15.7	23.1	8.9	10.5	6.1
1968	14.0	14.7	18.5	8.4	18.3	19.0	20.5	12.2	14.4	8.3
1969	13.2	12.2	23.0	8.8	18.9	15.5	23.5	10.1	11.0	8.6
1970	15.6	15.6	23.4	9.8	22.8	23.9	26.0	12.3	13.8	9.6
1971	15.4	14.6	25.4	9.4	23.1	20.8	29.9	11.8	13.1	9.4
1972	14.0	12.5	21.2	12.5	19.6	16.7	25.5	13.0	13.1	12.9
1973	17.6	16.8	29.0	11.2	27.0	25.2	31.9	13.5	14.8	11.2

information on such improvements is incomplete and, when available, represent national or republic totals—not the individual crop regions required for the crop prediction model.

15. As a surrogate for technological improvement, a time trend was tested in all crop regions and used where a high correlation between yields and time was found. In the districts where time was not correlated with crop yields (mainly the spring wheat belt of Western Siberia and Kazakhstan), new higher yielding varieties have not been introduced, and normal precipitation is inadequate for efficient use of most types of fertilizer.

Availability of Data

16. The weather data used in the prediction model include observations on precipitation and temperature for each of 27 crop regions from 1961 to 1973.[6] The observations are monthly averages except during the growing season, when average precipitation readings are available for ten-day periods. Weather data are not available for crop regions 28-33.

17. The yield data available cover all grain, spring wheat, winter wheat, winter rye, and spring barley for the years 1958-73. Yields were calculated from published data for the oblasts within a crop region, except for those few crop regions that coincide with the areas for which the USSR reports yields.

18. The weather data have their shortcomings. In the early years of collection—roughly 1961-65—data were transcribed by hand from weather maps, leading to errors in transcription. Only the most obvious of these errors could be corrected. More important, the observation for a given crop region is simply an unweighted average of the observations from all of the weather reporting stations in that region because of the lack of data on sown area below the oblast level. Thus, the weather observations in the major crop areas within a crop region were not accorded a proportionately greater weight.

19. Errors in the yield data can arise from inconsistencies in official Soviet statistics, the need to estimate unpublished data, and the need to estimate the ratio of sown area to aggregate yields. Although many inconsistencies appear in published Soviet sources, most involve a variance of only 0.1 centner per hectare. Wheat yields for the RSFSR are usually reported only for spring and winter wheat combined. By using additional information, yields for spring and winter wheat could be estimated separately. The estimates that could be checked were quite accurate; hence, any distortion is believed to be small. For some crops, such as spring barley, yield data were available for each division within a crop region but sown areas were unknown. The yield data therefore had to be aggregated by using estimated weights.[7]

The Prediction Model

Specification

20. The prediction model assumes that grain yield is a linear function of technology and weather. A linear time trend represents the influence of improved technology—for example, fallowing and increased use of fertilizers—and improved

[6] Although weather data have been collected since 1958, complete series for all variables used in the model are available only since 1961. The appendix describes the sources of weather and yield data. The 27 crop regions are identified on the map.

[7] These errors are discussed in more detail in the appendix.

seed varieties on grain yields. A linear combination of 11 weather variables comprises a weather index that represents the influence of weather on yields.

21. The Soviet grain yield prediction model is a set of 33 independent linear equations:

(1) $$Y^*_{ij} = c^*_{0j} + c^*_{1j} T_{ij} + c^*_{2j} V^*_{ij}$$

where Y^* = predicted yield,
T = time,
V^* = the weather index,
c^*_{0j}, c^*_{1j}, and c^*_{2j} are estimated model parameters,
i = year (1958 to 1974), and
j = crop region (1 to 33).

Estimates of sown area for the 33 crop regions are multiplied by the yield estimates for the crop regions to obtain forecasts of total production and the average yield. The estimates of sown area are based on Soviet press reports and past trends in each crop region.

22. The eleven weather variables used to form the weather index are:

- accumulated precipitation in millimeters from October through March;
- monthly precipitation in millimeters for April, May, June, July, and August; and
- average monthly temperatures in degrees centigrade for April, May, June, July, and August.

Because weather data are not available for crop regions 28 through 33, c_{2j} in equation (1) assumes the value of zero in these six crop regions. The weather variables listed above were selected from a set of 54 standard weather variables now available for weather-yield regression work on the USSR (see the appendix). In addition, new variables can be formed by combining standard variables.

Estimation

23. We estimated the Soviet grain yield prediction model in four steps. First, we estimated the effect of a time trend on yield in each of the 33 crop regions. We then estimated the influence of the weather variables, constructed the weather index, and finally estimated the parameters, c^*_{0j}, c^*_{1j}, and c^*_{2j} in equation (1).

24. In step one, we estimated the parameters a_{0j} and a_{1j} shown in equation (2). We derived these estimates by performing linear regressions for the 33 time-yield equations:

(2) $$Y_{ij} = a_{0j} + a_{1j} T_{ij} + e_{ij}$$

where Y_{ij} = reported yield,
i = year (1958 to 1973),
j = crop region (1 to 33), and
e_{ij} = an unexplained residual.

The calculated yields and residuals from equation (2) are

(3) $$Y^{**}_{ij} = a^*_{0j} + a^*_{1j} T^*_{ij}$$

(4) $$e^*_{ij} = Y_{ij} - Y^{**}_{ij}$$

The estimated time trend coefficient (a^*_{ij}) was not significantly different from zero for many eastern grain-growing regions. Therefore, the value of e^*_{ij} in equation (1) is set at zero for these regions.[8]

25. In the second step, we regressed the residuals calculated in equation (4) above on 11 weather variables to obtain an index of the influence of weather. The available weather data for each crop region are limited to 13 observations (1961-73) for crop regions 1-27. Therefore, in order to have enough observations to estimate the weather index, we arranged these 27 crop regions into 13 groups. We then computed the following regression equation:

$$(5) \qquad e^*_{ijm} = b_{om} + \sum_{k=1}^{11} b_{km} W_{ikjm} + u_{ijm}$$

where e^*_{ijm} = the estimated residual from equation (4) for year i in crop region j—which is assigned to one of the 13 regional groups (m),

W_{ikjm} = weather variable k (k= 1, . . . , 11), and

u_{ijm} = an unexplained residual.

26. The aggregation of the data into 13 groups of crop regions is called pooling the data.[9] The use of a single regression equation to express the relationship of yield to weather variables for a multi-region area implicitly assumes that the weather variables have the same influence on yield in each of the regions grouped. Thus, it was necessary to select groupings of crop regions with similar climates, soils, and cultivation practices.

27. In the third step, we calculated the weather index V^*_{ij} for each crop region. Given the parameters b^*_{km} (k=0, . . . , 11) estimated from equation (5), the weather index was computed as:

$$(6) \qquad V^*_{ij} = b^*_{om} + \sum_{k=1}^{11} b^*_{km} W_{kij}$$

where V^*_{ij} = weather index for crop region j in year i,

W_{kij} = weather variable k for crop region j in year i,

and all 27 crop regions (j) are assigned to one of 13 regional groups (m).

28. In the fourth step, we estimated the coefficients c_{oj}, c_{1j}, and c_{2j} (j=1, . . . , 33) in the model's grain yield prediction equation (1) by performing multivariate linear regression using equation (7) and the values for V^*_{ij} estimated in equation (6).

$$(7) \qquad Y_{ij} = c_{oj} + c_{1j} T_{ij} + c_{2j} V^*_{ij} + u_{ij}$$

The Pattern of the Regression Coefficients

29. The signs of the estimated coefficients of the weather regression—equation (5)—are shown in Table 2. The coefficients for October-March precipitation are generally positive (16 of 27 crop regions). Preseason precipitation would be expected to have a positive effect on yield, especially in the winter grain areas. Nonetheless, many of the negative coefficients are in winter grain

[8] Crop regions 6, 10, and 20-27.

[9] Pooling data is the use of cross-section data with time series data. Pooling the data increases the number of observations used in estimating the parameters in each equation and hence increases the number of degrees of freedom.

Table 2

Coefficient Signs of Weather Variables[1]

	Precipitation						Temperature				
Regions	Oct-Mar	Apr	May	Jun	Jul	Aug	Apr	May	Jun	Jul	Aug
1, 2	+	+		+							
3	+	+	−	N.A.	N.A.	N.A.	+	+	N.A.	N.A.	N.A.
4, 5, 6			+	+			+	−	+		−
7, 8		+	+	+	−	−	+				−
9, 10	+	−	+	+		+	+	+	−	+	−
11, 12	+	+	−	+	−	+	+		+	−	−
13, 14	−	+	−	+	+		+	−	−	−	−
15, 18	−	−	−	+	+	−	+	−	+	−	−
16, 19	+	+	+	+	+	+	+	+	−	−	−
17	+	−	+	N.A.	N.A.	N.A.	+	+	N.A.	N.A.	N.A.
20, 26, 27	+	+	+	+	−	−	+	−	−	−	+
21, 22, 23	+	+	+	−	+	+	−	−	−	−	+
24, 25	−	+	+	+	+	+	+	−	+	−	+

[1] These coefficients resulted from using the 11 weather variables enumerated with the grouping of crop regions given. Regions 3 and 17 were estimated separately with a reduced set of weather variables.

areas. Precipitation seems to be important in April, May. and June, with the coefficients generally positive. The coefficients for April temperature are mostly positive, but several turn negative in May and June, suggesting that very hot weather at the onset of heading is not desirable. The need for dry weather at harvest time clearly shows up in the winter grain areas in July, as most precipitation coefficients shift from positive to negative. The spring grain areas retain their positive coefficients in July, but two of them become negative in August.

30. The results presented in Table 2 are generally encouraging, although the signs do not correspond with agronomic theory in several instances. Discrepancies could arise for a number of reasons:

- Although weather and time were assumed to be unrelated, in fact weather may be cyclical. Since 1961, precipitation in the USSR has tended to increase and average temperature has declined slightly, affecting the estimated time trend in equation (2) and biasing the weather variable coefficients.
- Weather variables also may be related within the same year, contrary to the model's assumption that the weather variables in equation (5) are independent. Colinearity of this nature could easily cause a sign reversal in the weather coefficients.
- Several of the weather variable coefficients were statistically insignificant.[10] There are three likely causes: (1) a variable truly is not related to yield, (2) a variable is colinear with another weather variable and therefore does not receive its full credit,[11] and (3) the small number of observations precludes an accurate test of significance.

[10] A coefficient is statistically insignificant if it cannot be stated with a high probability (usually 95%) that the true value of the coefficient is non-zero.

[11] A number of tests were made by reestimating equation (5) after excluding one or more insignificant weather variables. These tests did not show any important sign reversals, but did sharply improve the significance of the remaining variables. This indicates the presence of colinearity among the weather variables.

Approved For Release 2000/09/14 : CIA-RDP86T00608R000500230008-5

The types of grain grown within each crop region and in the nation change over time because of weather, shifts in demand, and technological change. The coefficients of the weather equation reflect changes in the structure of the sown area and therefore are unlikely to apply fully in any particular year.

PREDICTIONS FOR 1974

31. The use of the model to forecast grain yields began in April and continued throughout the summer months. The forecasts were updated every ten days, as new weather data were received. Two basic types of forecasts were made during the 1974 crop year. The first used only weather variables for which 1974 data had been received. For example, the forecast made as of the end of April did not include the influence of weather in May, June, July, and August. The second type used all weather variables, substituting long-run norm values for variables not yet reported. The results reported here are all from the second type of forecast.

32. In 1974 the estimate of grain production ranged from 184 to 203 million tons. A preseason estimate of 185.9 millions tons was based on longrun norm values for all weather variables. The tabulation below shows how the predictions changed as additional weather data were received.

Weather Through	Predicted Yield (Centners per Hectare)	Predicted 1974 Grain Production (Million Tons)
Preseason	14.5	185.9
March	14.6	186.3
April	14.4	184.2
May	14.8	189.9
June	15.9	203.3
July	15.5	197.8
August	15.6	199.4

The development of the harvest prediction becomes clearer when republic estimates are examined. The gains made in Kazakhstan through April were more than canceled out by poor weather in the rest of the country. In May, Kazakhstan fared poorly, but developments in the Ukraine easily compensated for Kazakhstan's diminished prospects and provided the first hope for a good year. In June, predicted yields in the Ukraine continued to increase, and those in the RSFSR made a very large jump as a result of abundant rains. Predicted harvests in all areas fell off in July and recovered slightly in August. The final line in the tabulation of predicted 1974 grain yields below shows the production as reported in December 1974. The predicted values for the RSFSR and the Ukraine are 3.7 million and 1.9 million tons too high. Since this is more than the difference of 3.9 million tons for the USSR prediction, the prediction for Kazakhstan and other areas must be too low.

Weather Through	Millions Tons				
	RSFSR	Ukraine	Kazakhstan	Other Areas	Total
Preseason	109.4	39.6	19.8	17.1	185.9
March	109.2	40.3	19.8	17.0	186.3
April	107.8	39.1	21.8	15.5	184.2
May	108.8	46.0	19.1	16.0	189.9
June	116.3	48.6	21.0	17.4	203.3
July	114.0	47.1	20.5	16.2	197.8
August	115.3	47.9	19.5	16.7	199.4
Reported production	111.6	46.0	N.A.	N.A.	195.5

33. Tabulations of the progress of the predictions for winter and spring wheat indicate that winter wheat caused the increase in the predicted grain crop in May, and spring wheat caused the large jump in June.

WINTER WHEAT

	Million Tons			
Weather Through	RSFSR	Ukraine	Other Areas	Total
Preseason	16.2	20.1	5.6	41.9
March	15.7	20.1	5.6	4[illegible]
April	14.3	20.3	5.4	40.0
May	15.8	22.9	5.7	44.4
June	16.3	23.4	5.9	45.6
July	14.4	22.9	5.7	43.0

SPRING WHEAT

	Million Tons			
	RSFSR	Kazakhstan	Other Areas	Total
Preseason	32.1	12.3	0.1	44.5
March	32.2	12.4	0.1	44.7
April	32.6	13.1	0.1	45.8
May	32.4	11.6	0.1	44.1
June	36.2	12.6	0.1	48.9
July	35.9	11.6	0.1	47.6
August	36.6	10.6	0.1	47.3

EVALUATION OF THE MODEL

34. The value of the model was tested by computing its prediction errors using historical data and by comparing its performance with other models. First, we deleted data and reestimated the model coefficients from the remaining data.[12] We predicted crop yields from equation (1) for each of the 33 crop regions. Next we computed an average USSR yield using reported sown areas. We then computed the difference between our predicted USSR yield and the yield reported by the USSR Central Statistical Administration. The following tabulation shows the direction and percentage magnitude of the prediction errors, i.e.,

$$\frac{(\text{predicted yield} - \text{reported yield})}{\text{reported yield}} \times 100$$

Years Covered in Model Estimation	Percentage Prediction Errors for All Grain Yields						
	1968	1969	1970	1971	1972	1973	1974 [1]
1958-67	−10.0	4.5	−7.2	−11.8	−1.6	−13.3	−10.1
1958-68	—	−6.3	0.4	−13.2	−1.6	−14.5	−5.9
1958-69	—	—	3.4	−9.0	6.2	−10.0	−4.2
1958-70	—	—	—	−8.6	3.8	−7.6	−0.8
1958-71	—	—	—	—	8.2	−9.2	2.5
1958-72	—	—	—	—	—	−8.1	1.2
1958-73	—	—	—	—	—	—	2.0

[1] Total sown area has not been reported for 1974; the estimated yield is based on an estimated sown area of 128 million hectares.

---

[12] This procedure resembles the way the model would be used in practice. As data for additional years become available, the model would be reestimated.

This test suggests that an error of between 2% and 8% should be expected—the range of the precentage errors shown on the diagonal for the years 1969-74. While the upper part of this range is too high from the standpoint of making useful forecasts,[13] experience has shown that collateral information can be used to detect anomalies and reduce the error.

35. Next we constructed several naive models to determine whether our model provides better predictions. Even the best of these naive models did not perform well enough to merit consideration. In the best of the naive models, yield was a linear function of time, past yields, and selected weather variables. The data for all crop regions for which weather data were available (regions 1-27) were used in the same equation. This equation was able to describe historical yields fairly well (the coefficient of determination was 0.8), but it did not capture the cyclical behavior of Soviet agriculture, and the average absolute percentage prediction error was almost double that of our disaggregated model, based on predictions for 1969-72.

36. In evaluating the model described in this publication (and especially the predictions for 1974), the fact that production forecasts depend on the accuracy of sown area must be kept in mind. Although actual sown areas have not been reported for 1974, the errors in estimates of total sown areas in 1973 ranged from 2.1% for all grain to 13.1% for winter wheat, as shown below.

Crop	Million Hectares: Actual	Million Hectares: Estimates Used in Prediction Model	Percentage Error
All grain	126.7	124.0	− 2.1
Spring wheat	44.8	47.0	4.9
Winter wheat	18.3	15.9	−13.1

Sown area data by crop region in 1973 have been received only for the Ukraine. A comparison of actual sown areas in the Ukraine in 1973 with those estimated on the basis of laborious review of published information also shows substantial errors, as follows:

Crop Region	All Grain: Million Hectares: Actual	All Grain: Million Hectares: Estimated	All Grain: Percentage Error [1]	Winter Wheat: Million Hectares: Actual	Winter Wheat: Million Hectares: Estimated	Winter Wheat: Percentage Error [1]
Total Ukraine [2]	16.6	15.2	− 8.8	8.3	7.0	−16.4
Western Ukraine	2.5	2.3	− 6.9	1.2	1.2	− 3.3
North Central Ukraine	4.5	4.2	− 6.6	2.1	2.0	− 3.9
Northeast Ukraine	2.6	2.4	− 7.4	1.0	1.2	14.9
Eastern Ukraine	3.6	3.3	− 9.2	1.9	1.0	−47.9
Southern Ukraine	3.5	3.0	−13.5	2.2	1.7	−22.8

[1] Percentages were derived from unrounded data.

[2] Because of rounding, components may not add to the totals shown.

Knowledge of the actual distribution of the 1973 Ukrainian sown area would not have greatly improved the estimate of the overall Ukrainian yield (from 28.76 to 28.82 centners per hectare for winter wheat), but the error in the predicted total harvest would have been sharply reduced (from −24.8% to −9.7% for winter wheat). Thus, the effect on crop estimates of errors in the estimates of the regional distribution of sown area appear to be much less than the effect of errors in estimates of the total sown area.

[13] An error of 8% applied to a national yield of 15 centners per hectare and a sown area of 128 million hectares, results in a harvest error of 15 million tons.

37. Although the model described in this publication presumably can be improved as data on weather and yields accumulate, it must be used with care. The essence of the model is a set of equations which estimate a linear relationship between the yield of a given crop and a number of weather variables and a time trend. There are several problems inherent in the model:

- The time trend is used as a surrogate for factors that the model, because of lack of data, cannot account for specifically.
- In any event, since the time trend averages the long-term effect of other variables, it will not reflect their influence accurately in any given year. The impact of a sharp increase in the supply of fertilizer in a given year, for example, would not be reflected in the results of the model.
- The weather indexes measure the influence of weather variables in an "average" year, but each year is different with respect to the timing of crop developments.
- The assumed linear influence of the weather variables undoubtedly fails to capture the effects of extreme weather accurately.

38. The range of the expected prediction error can probably be narrowed by introducing nonlinear variables, additional experiments with the grouping of crop regions, or other weather variables. Also, the data base will gradually increase, adding to the number of observations available to estimate the parameters of the weather variable. Despite these improvements, it will still be necessary to consider collateral information about variations in the crop calendar, harvesting conditions, use of fertilizer, and amounts and quality of seed available and to adjust the model predictions accordingly.

APPENDIX

DESCRIPTION OF THE DATA BASE

Weather Data

Four basic types of weather data covering the period from 1961 to the present are currently maintained on computer files for use in weather-yield regressions—precipitation by decade [1] for the months of April through September, total precipitation by month, average monthly temperature, and soil moisture as of the last day of each month.

Sources and Reliability

Basic precipitation and temperature data for individual weather stations in the USSR are broadcast two to four times daily by Soviet radio stations.[2] From these broadcasts, weather data are interpolated for some 134 grid points spaced roughly 75 miles apart in a square pattern throughout the area studied. Periodically, total precipitation, mean temperature, and soil moisture are calculated for each grid point and aggregated into values for crop districts.

Reliability of the weather data is primarily a function of the density of coverage—that is, the number of individual weather stations from which reports emanate affects the interpolated data for a given grid point. The density of coverage is most critical for the precipitation values because there can be considerable variation in precipitation over a relatively small area, especially in hilly country. Temperature values, on the other hand, are considered to be more continuous.

Before April 1965 the density of coverage averaged one reporting weather station per grid point, largely because of limitations imposed by the time-consuming manual method of interpolation used. Since that date, computer analysis has applied a complex method of interpolation that permitted increasing the density of coverage to four or five reporting weather stations per grid point. Consequently, the interpolated values for precipitation and temperature since April 1965 are considerably better indicators of weather conditions at the grid points than those for earlier years.

Storage and Handling of Weather Data

The basic weather data are received at approximately ten-day intervals and stored on a computer disk file. Six categories of weather variables are stored in the file in the following order:

1. First decade precipitation
2. Second decade precipitation
3. Third decade precipitation
4. Total monthly precipitation
5. Average monthly temperature
6. Soil moisture (last day of month)

[1] Each month is divided into three periods called "decades."

[2] As a member of the World Meteorological Organization, the USSR shares such information with foreign countries.

Each of the three decade categories contains only six variables (representing the months of September and April through August) while the remaining three categories contain 12 variables each (one for each month of the year). A complete list of the weather variables is shown in Table A-1.

TABLE A-1

STANDARD WEATHER VARIABLES AVAILABLE

Number	Name of Variables
1	Precipitation, 1st Decade of September
2	Precipitation, 1st Decade of April
3	Precipitation, 1st Decade of May
4	Precipitation, 1st Decade of June
5	Precipitation, 1st Decade of July
6	Precipitation, 1st Decade of August
7	Precipitation, 2d Decade of September
8	Precipitation, 2d Decade of April
9	Precipitation, 2d Decade of May
10	Precipitation, 2d Decade of June
11	Precipitation, 2d Decade of July
12	Precipitation, 2d Decade of August
13	Precipitation, 3d Decade of September
14	Precipitation, 3d Decade of April
15	Precipitation, 3d Decade of May
16	Precipitation, 3d Decade of June
17	Precipitation, 3d Decade of July
18	Precipitation, 3d Decade of August
19	Precipitation—September
20	Precipitation—October
21	Precipitation—November
22	Precipitation—December
23	Precipitation—January
24	Precipitation—February
25	Precipitation—March
26	Precipitation—April
27	Precipitation—May
28	Precipitation—June
29	Precipitation—July
30	Precipitation—August
31	Average Temperature (Absolute Value)—September
32	Average Temperature (Absolute Value)—October
33	Average Temperature (Absolute Value)—November
34	Average Temperature (Absolute Value)—December
35	Average Temperature (Absolute Value)—January
36	Average Temperature (Absolute Value)—February
37	Average Temperature (Absolute Value)—March
38	Average Temperature (Absolute Value)—April
39	Average Temperature (Absolute Value)—May
40	Average Temperature (Absolute Value)—June
41	Average Temperature (Absolute Value)—July
42	Average Temperature (Absolute Value)—August
43	Soil Moisture (Last Day of the Month)—September
44	Soil Moisture (Last Day of the Month)—October
45	Soil Moisture (Last Day of the Month)—November
46	Soil Moisture (Last Day of the Month)—December
47	Soil Moisture (Last Day of the Month)—January
48	Soil Moisture (Last Day of the Month)—February
49	Soil Moisture (Last Day of the Month)—March
50	Soil Moisture (Last Day of the Month)—April
51	Soil Moisture (Last Day of the Month)—May
52	Soil Moisture (Last Day of the Month)—June
53	Soil Moisture (Last Day of the Month)—July
54	Soil Moisture (Last Day of the Month)—August

Crop Data

The data bases for harvest and sown area have been assembled for five grain crops: (1) all grain, (2) winter wheat, (3) spring wheat, (4) spring barley, and (5) winter rye. These data bases are computerized and will be referred to as a "file." Each file has a common structure; all are processed by the same computer program. The data sources are not the same, however, and some data have been estimated.

Data Required

For each crop, data are required for all 33 crop regions. The 33 crop regions are composed of 127 administrative divisions—12 Soviet Socialist Republics (SSRs), 16 Autonomous Soviet Socialist Republics (ASSRs), 93 oblasts, and 6 krays. For each of the 127 divisions, data on the size of the harvested crop in thousands of tons and the harvested area in thousands of hectares are needed. The computer program computes yields for each of the 127 divisions in centners per hectare, sums the harvested crop and the harvested area for all of the divisions within a crop region, and computes the yield for each crop region.

Data Sources

All Grain: Data on production and sown area for all grain are usually available from Soviet statistical handbooks published for the area and year of interest. Although data for Kazakhstan in 1962-64 and 1969-70 are missing, five-year moving averages for spring wheat in selected Kazakhstan oblasts were recently published. Since spring wheat is dominant in these areas, these data were used to estimate yields for 1962-64 and 1969-70 on th assumption that spring wheat yields were the same as for all grain.

No other all grain data are estimated except in the sense that the computer program computes the yields. Sometimes this causes a variation of 0.1 centner per hectare from the yields published in Soviet statistical handbooks.

Spring and Winter Wheat: Data for Estonia, Latvia, Lithuania, Belorussia, Moldavia, Kazakhstan, and the Ukraine are taken from various statistical abstracts. The statistical handbooks for the RSFSR normally publish data for all wheat only. In order to estimate data for winter and spring wheat separately, additional sources and estimation methods had to be used. Several estimation procedures were used, depending on the data available.

The best estimates are for the years 1960, 1962, and 1963. The only data missing are the harvested crop by administrative division, which can be estimated by,

$$(1) \qquad \text{Harvest} = \text{Area} \times \text{Yield}$$

for each division. To adjust for errors introduced by the use of rounded data, published data on spring and winter wheat by economic region were used. The actual harvested crop for an economic region should equal the sum of the calculated harvested crops for the divisions within that region, and the published data on total wheat for each division should equal the sum of the calculated winter wheat harvest and the calculated spring wheat harvest. The calculated harvest data were adjusted to meet these conditions while remaining within the range of possible rounding error.

The estimates for 1958 and 1971 use essentially the same method. For 1958, exact data on the harvested crop by economic region are not available. It was

estimated using data on area and yields by economic region. Hence the adjustments to the calculated harvest are not as accurate. For 1971, data on the harvested area of winter and spring wheat are not available by economic region or by administrative division. Data on the harvested crop are available; hence equation (1) can be rearranged as

(2) $$\text{Area} = \text{Harvest} \div \text{Yield}$$

in order to estimate sown areas. The estimated sown areas were then refined using the same type of check as described above.

For 1965-69, data for spring or winter wheat have not been reported by administrative division, for the harvested crop or the harvested area. Equations (3) and (4) were solved simultaneously to estimate these data for each administrative division.

(3) $$H = (YS) \times (AS) + (YW) \times (AW)$$

(4) $$A = AS + AW$$

where:

H = total wheat harvest
YS = yield of spring wheat
YW = yield of winter wheat
AS = sown area of spring wheat
AW = sown area of winter wheat
A = total sown area of wheat

H, A, YS, and YW are known. The solution is:

(5) $$AW = \frac{H - (A) \times (YS)}{YW - YS}$$

(6) $$AS = A - AW$$

Because the denominator of equation (5) is the difference of two numbers of similar magnitudes, these estimates are susceptible to large errors. The refinement procedures discussed above were also used for these calculations, but had a much greater impact because larger errors were encountered.

In order to gauge the accuracy of this method, equations (5) and (6) were computed for all years 1958-69. Direct comparisons with published data for 1958, 1960, 1962, and 1963 showed that as long as the difference between spring and winter wheat yields—the denominator of (5)—was not close to zero, the estimates were quite good. The refinements using control totals for economic regions and the relative stability over time of sown areas provide additional confidence in the estimates.

The situation for 1961 and 1964 is similar to that for 1965-69, except that control totals for harvested area by economic region are not available. But the harvested crop and yield together imply an estimated harvested area for economic regions that is sufficiently accurate to provide usable refinements of estimates for administrative divisions. For 1959, there are data by economic region for harvested area but not for harvested crops. As for 1961 and 1964, crops can be estimated using data on yields and harvested area by economic region.

Data for 1970 were the hardest to estimate. Data for winter and spring wheat have not been reported separately by oblast, only for total wheat. But the source of data for 1971 also gives 1971 results as a percent of the 1966-70 average. From

this information, the average yields and average harvested crops could be calculated for 1966-70. The estimates for 1970 are then given by

$$(7) \qquad Y_{70} = 5Y - \sum_{1966}^{1969} Y_i$$

$$A_{70} = 5A - \sum_{1966}^{1969} A_i$$

where:

Y_i = yield in year i
A_i = sown area in year i
Y = average yield in 1966-70
A = average sown area in 1966-70

These estimates were then refined by checking them against economic region data and against data for total wheat by administrative division. In most cases the estimates did not have to be substantially revised.

Spring Barley: Data on spring barley are not published consistently. The statistical handbooks for the Ukraine regularly publish data by oblast, but other handbooks normally do not. The data for the Baltic districts, Belorussia, and Moldavia come mostly from the agricultural handbook *Sel'skoye Khozyaystvo* (1970) and are directly available as yields in centners per hectare. Data for Kazakhstan are not available for the years 1962-64 and 1969-71.

Data for the RSFSR are available at the administrative division level by yield only for the years 1958-69. Weights were estimated in order to combine the yields of the divisions within a crop region. They are based on data on sown area derived from a Soviet journal article and from a Soviet agricultural atlas.[3] The only data available for 1970 and 1971 are for economic regions. They are usable for two regions, the Central and the Volga-Vyatsk, which are very close in definition to crop districts 13 and 14. The weighted averages for each crop district were computed and the results entered in the place of one of the divisions in each crop district.

Winter Rye: Data by oblast for the RSFSR, which has 75% of the area sown to winter rye, have been published only through 1969. Oblast data for the Ukraine (less than 10% of the total area sown to winter rye) are available only since 1965. Reasonable estimates for 1958-64 can be made at the crop region level, however, from the data for economic regions which are published annually. Various republic handbooks and the USSR handbook supply the data for the remaining crop regions.

---

[3] *Zernovoye khozyaystvo*, No. 3, 1972, p. 17-20 and *Atlas sel'skogo khozyaystva SSSR*, Moscow, 1960.

Approved For Release 2000/09/14 : CIA-RDP86T00608R000500230008-5